Electronic Version
Stylesheet Version v1.1.1

# Claims

[c1] 1. A method for processing digital messages on an electronic communication system, each message having a header and a body, comprising: identifying a set of characteristics of a first message, the set including; addresses extracted from the header and body of the message; anda condensed representation of the message body produced by: eliminating message content not perceptible in the normal display mode of the message; converting the perceptible message content to a standardized format characterized by limited degeneracy; generating a plurality of hash values which represents the converted content of the message body; storing the set of identified characteristics of the first message in a first bulk message envelope, the first bulk message envelope including a frequency index; identifying the same set of characteristics of a second message; comparing the set of identified characteristics of the second message to the first bulk message envelope; upon determining that the second message has characteristics dissimilar to those of the first bulk message envelope, storing the set of identified characteristics of the second message in a second bulk message envelope, the second

bulk message envelope including a frequency index with a unitary value; upon determining that the second message has characteristics similar to those of the first bulk message envelope, increasing the frequency index of the first bulk message envelope by a unitary increment.

[c2] 2. The method of claim 1, further comprising:
identifying the same set of characteristics of a third message;
comparing the set of identified characteristics of the third message to the first and second bulk message envelopes;
upon determining that the third message has characteristics similar to those of either the first or second bulk message envelopes, increasing the frequency index of the most similar bulk message envelope by a unitary increment;
upon determining that the third message has characteristics dissimilar to those of the first and second bulk message envelopes, storing the set of identified characteristics of the third message in a third bulk message envelope, the third bulk message envelope including a frequency index with a unitary value.

[c3] 3. The method of claim 1 in which the step of identifying the characteristics of each message comprises:
transforming the message to a reduced format; process-

ing the reduced message to derive a condensed representation of the reduced message.

[c4] 4. The method of claim 3 in which the condensed representation comprises plural hashes.

[c5] 5. The method of claim 3 in which the step of transforming the body to a reduced format comprises:
eliminating non-communicative information from the message.

[c6] 6. The method of claim 3 in which the step of translating the body to a reduced format comprises:
conforming communicative information in the message to a standardized format characterized by limited redundancy.

[c7] 7. The method of claim 3 in which the step of translating the body to a reduced format comprises:
eliminating address information from the message.

[c8] 8. The method of claim 1 in which a set of characteristics of each message comprises:
address data associated with the message.

[c9] 9. The method of claim 8 in which the address data comprises:
the purported originating address of the message.

[c10] 10. The method of claim 8 in which the address data comprises:
one or more addresses included in the body of the message.

[c11] 11. The method of claim 8 in which the address data comprises:
one or more addresses through which the message has purportedly been relayed.

[c12] 12. An electronic communication system comprising interconnected entities for transmission and receipt of messages, the system comprising, in at least one of said entities, a subsystem for processing messages comprising:
a unit which identifies a set of characteristics of a message; a memory which stores the set of identified characteristics of messages in a plurality of bulk message envelopes, each bulk message envelope including a frequency index; a unit which compares the set of identified characteristics of a message to the bulk message envelopes, and if the identified characteristics are similar to a stored bulk message envelope, increasing the frequency index of the bulk message envelope in response, and if the identified characteristics are dissimilar to any stored bulk message envelope, causing the set of identi-

fied characteristics to be stored in the memory as an additional bulk message envelope.

[c13] 13. A computer program embodied on a computer-readable medium and/or memory device for providing a subsystem for processing messages comprising:
an identification segment for extracting a set of characteristics of a message; a storage segment for storing the set of identified characteristics of a message in a bulk message envelope, each bulk message envelope including a frequency index; a comparison segment for comparing the set of identified characteristics of a message to the bulk message envelopes, and if the identified characteristics are similar to a stored bulk message envelope, increasing the frequency index of the bulk message envelope by a unitary increment, and if the identified characteristics are dissimilar to any stored bulk message envelope, causing the set of identified characteristics to be stored in the memory as an additional bulk message envelope having a frequency index with a unitary value.

[c14] 14. An article of manufacture comprising:
a machine readable medium and/or memory device that provides instructions that, if executed by a machine operatively connected to an electronic messaging system, will cause the machine to perform operations including:

identifying a set of characteristics of a first message; storing the set of identified characteristics of the first message in a first bulk message envelope, the first bulk message envelope including a frequency index; identifying the same set of characteristics of a second message; comparing the set of identified characteristics of the second message to the first bulk message envelope; upon determining that the second message has characteristics similar to those of the first bulk message envelope, increasing the frequency index of the first bulk message envelope by a unitary value; upon determining that the second message has characteristics dissimilar to those of the first bulk message envelope, storing the set of identified characteristics of the second message in a second bulk message envelope, the second bulk message envelope including a frequency index with a unitary value.

[c15] 15. The method of claim 3 in which a set of characteristics of a bulk message envelope include heuristic properties relating to the formatting or presentation of the data in the messages.

[c16] 16. The method of claim 3 in which from a set of bulk message envelopes thusly made, and using various characteristics of these, a real time black list (RBL) of spammer domains are derived.

[c17] 17. The method of claim 16 in which the RBL is applied against the current set of messages, possibly with a delay to detect and block current types of unwanted messages, or against a new set of messages, to block unwanted messages.

[c18] 18. QuickMarkQuickMarkThe method of claim 16 in which the RBL is used by various routing services or gateways or relay machines to block communications with entries in the RBL.

[c19] 19. The method of claim 3 in which a user can maintain a "gray list" of desired bulk message senders, and the user's message provider uses claim 3 to find bulk messages addressed to the user, and from these bulk messages, forwards only those from senders on the gray list, to the user, where the determination of the sender of a message may involve examining the contents of a message, in addition to examining the purported sender field and other entries in the header.

[c20] 20. The method of claim 4 in which these plural hashes may be exchanged by different organizations or users to detect messages seen by others, in an anonymous query manner that preserves the privacy of the original messages.

[c21] 21. The method of claim 4 in which these plural hashes may be found in an adaptive hashing manner.

[c22] 22. The method of claim 3 in which the subsets of the bulk message envelopes are chosen, to which further filtering is applied; where the filtering might include Bayesian, neural network or other techniques, some of which are possibly dependent on human languages; where the subsets may be derived using various values of the bulk message envelopes, including, but not limited to, the frequency of each envelope.

[c23] 23. The method of claim 3 in which the bulk message envelopes found from messages in one electronic communication space are used to compare and correlate with those derived from messages or data in another electronic communication space.